心理科学进展 2023, Vol. 31, No. 11, 1–14
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2023.001

机器学习在发展性阅读障碍儿童早期筛查中的应用*

卜晓鸥 王 耀 杜亚雯 王 沛
(华东师范大学教育学部特殊教育学系, 上海 200062)

摘 要 发展性阅读障碍严重影响儿童的学业成就、心理健康和社会适应能力。近年来, 机器学习因其强大的数据处理和挖掘能力逐渐被应用到阅读障碍儿童的早期筛查中, 在标准化心理教育测试、眼动追踪、游戏测试、脑成像等多个领域积累了较为丰富的成果, 获得了更加精准高效、灵活可靠的分类结果。然而, 机器学习在对象选取、数据采集、转化潜力和安全隐私等方面仍存在局限性。未来研究需要重点关注学龄前阅读障碍儿童的早期筛查系统的科学性, 同时积极构建多模态数据库、在多种算法中寻找最佳算法以获取最优参数, 最终实现临床实践中的广泛使用。

关键词 发展性阅读障碍, 机器学习, 早期筛查, 儿童
分类号 R395

1 引言

发展性阅读障碍(Developmental dyslexia, DD)是一种极其复杂的神经发育性障碍, 其核心特征是尽管个体的智力正常, 视、听觉功能完好, 但是仍然表现出持续的阅读、拼写和写作困难(Kaisar, 2020)。阅读障碍在不同的语言和文化中的发生率约为 5%~15% (Tamboer et al., 2016), 并且存在代际传递现象(Zahia et al., 2020)。目前, 儿童通常在 2 年级或更高年级在掌握阅读技能的过程中才有可能被识别出存在阅读障碍(Sanfilippo et al., 2020)。在经济发展落后的国家, 贫困儿童发现存在阅读障碍的年龄更晚(Ballester et al., 2021)。此时往往已经错过了最佳的干预窗口期, 即幼儿园至 1 年级大脑可塑性增强的早期阶段(Fox et al., 2010)。大量研究已然发现, 患有阅读障碍的儿童会深陷学习成绩低下、自我效能感降低和学习动力不足的恶性循环中(Burns et al., 2022), 甚至出现极高的辍学率和心理健康问题(Livingston et al., 2018)。如果此类儿童未能得到及时的识别和干预, 阅读障碍的负面影响可能会从童年早期一直持续至成年期(Farah et al., 2021)。因此, 进行高效的早期筛查, 提供有效的早期干预, 对于阅读障碍儿童的发展具有关键性意义。

迄今为止, 阅读障碍的筛查主要借助于标准化心理教育测试(Lee et al., 2022)、眼动追踪(Prabha & Bhargavi, 2019)、网络/手机游戏(Borleffs et al., 2018)以及脑成像技术(Usman et al., 2021)等手段。标准化心理教育测试通常采用智商−成就差异模式(IQ−achievement discrepancy) (Fletcher et al., 2019)、干预−应答模式(response to intervention, RTI) (Miciak et al., 2014)、优势与弱势模式(pattern of strengths and weaknesses, PSW) (Hale et al., 2010)来评估和量化个体的智力、语音加工、阅读技能和词汇发展等认知能力, 进而达到识别阅读障碍者的目的(Miciak & Fletcher, 2020)。就眼动追踪技术的应用而言, 研究者通过记录阅读过程中的眼动特征来区分阅读障碍儿童和非阅读障碍儿童, 这些特征包括注视/回视时间和次数、眼跳幅度和次数、眨眼频率和次数以及双眼协调性等(Hmimdi et al., 2021)。也有研究者以游戏化的形式生成具体的语音测试或认知测试, 开发基于网络技术的电子学习系统和手机游戏(例如, Deslixate 和 GraphoGame), 旨在通过教育游戏识别阅读障碍儿童(Larco et al., 2021; Ojanen et al., 2015)。随着认知神经技术的发展, 越来越多的研究使用脑

收稿日期: 2022-11-08
* "华东师范大学幸福之花'音乐画的脑智机制及促进儿童艺术教育发展的实践进路'"资助。
通信作者: 王沛, E-mail: wangpei1970@163.com

成像技术获取大脑的结构、形态、功能激活和几何特性，利用组间均值差异来区分阅读障碍儿童和典型发展儿童(Livingston et al., 2018; Sihvonen et al., 2021; Yang et al., 2021)。

然而，阅读障碍儿童的症状具有巨大的个体差异性，比如不同的病源因素会导致不同的阅读障碍亚类型(Aaron et al., 1999)。加之传统的阅读障碍检测技术低效耗时，敏感性和特异性指标不明确，难以满足大规模并快速筛查阅读障碍儿童的需求(Usman et al., 2021)。更重要的是，阅读障碍与多种神经、行为和环境因素有关，这些因素以复杂的方式相互作用导致了阅读障碍(Catts et al., 2017; McGrath et al., 2020)。因此，仅凭单一因素或少数因素结合无法完成对阅读障碍患者的精确诊断(Catts & Petscher, 2022)，即使是传统的多因素方法也无法涵盖所有可能的因素和关系(Walda et al., 2022)。研究复杂系统的一种相对新颖有效的方法是机器学习(Kaisar, 2020)。机器学习(Machine Learning, ML)是使用计算机算法让机器从大量经验数据中学习规律，自动识别模式以做出预测或决策(Gilvary et al., 2020)。近年来，因其能够提供更高的检测精度和更好的预测结果，一些研究者们开始尝试应用机器学习来提高阅读障碍的筛查的精度与敏感性。为此，本研究通过整合机器学习在阅读障碍筛查中的最新进展、主要应用范围、未来可能的发展方向，旨在廓清阅读障碍的机器学习研究可能的发展路径与发展思路。

2 方法

我们对 2016 年以来用于分类和识别阅读障碍的机器学习方法的研究进行文献搜索，使用的数据库包括 Web of Science、Elsevier Science Direct、EBSCO 和 PubMed。检索关键词为“Dyslexia/Reading Disability” AND “Identification/Screening/Detection/Recognition/Prediction/Diagnosis” AND “Machine Learning/Deep Learning/Artificial Intelligence” AND “Child/Children/ Preschool”。考虑到机器学习技术的飞速发展和迭代，并且第一篇中文阅读障碍的机器学习研究发表于 2016 年，因此文献检索的日期范围设定为 2016 年 1 月 1 日~2022 年 10 月 1 日。文献纳入标准为: (1)文献为英文实证期刊论文和会议论文，全文可得并包含明确的研究问题、方法和结论，研究结论有翔实的数据支撑；(2)研究对象为 18 岁以下的儿童，设置典型发展对照组和阅读障碍组。阅读障碍儿童无其他共病(如计算障碍、书写障碍、自闭症等)。(3)文献使用/组合使用机器学习方法筛查阅读障碍。我们依据上述标准进行独立筛查，最后确定纳入本次系统综述的文献数量为 25 篇(见表 1)。图 1 和图 2 分别展示了文献筛选流程和文献检索完成后的文献年度分布情况。

3 基于机器学习的阅读障碍早期筛查的主要步骤

3.1 数据采集

基于机器学习的阅读障碍筛查的第一步是使

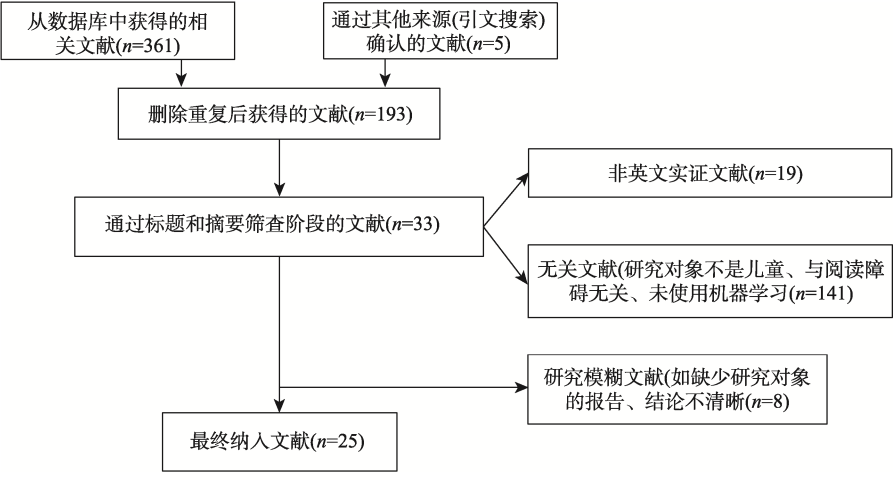


图 1 文献筛选流程图

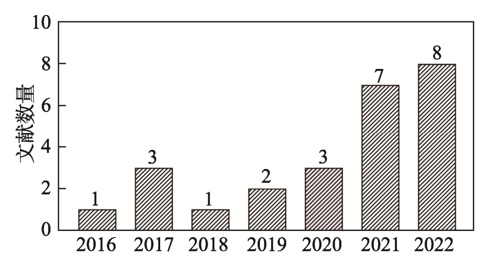


图 2　文献检索完成后的文献年度分布

用相应的技术手段获取数据。

标准化心理教育测试为机器学习的模型构建提供了最早、最广泛的证据。其所提供的数据显示了阅读障碍患者明显的行为表现, 主要包括阅读、语音加工、工作记忆、视听辨别等。Chen 等人(2017)使用荷兰版的 McArthur-Bates 沟通发展量表(N-CDI)测量了 476 名 17~35 个月的典型发展儿童的早期词汇发展能力, 使用机器学习算法预测具有阅读障碍家庭风险的儿童。Shamir 等人(2019)采用自行开发的阅读障碍简短筛查工具(Zippy 6)测量了 125 名儿童(6~14 岁, 其中阅读障碍儿童 81 名)的认知能力和语音能力, 并使用机器学习算法区分阅读障碍儿童和典型发展儿童。Tolami 等人(2021)收集了 54 名 8~11 岁儿童(29 名阅读障碍儿童)的语言样本, 使用计算语言学方法提取拼写和语法错误、词汇多样性、语法复杂性指数和可读性等阅读障碍的差异特征, 利用机器学习模型诊断阅读障碍。在中文阅读障碍研究中, Wang 和 Bi (2022)收集了 399 名 7~13 岁阅读障碍儿童的认知测试集, 在测量阅读流畅性、阅读准确率、语音意识、语素意识、快速命名和正字法意识的基础上, 使用深度学习模型预测中文阅读障碍儿童的症候。Lee 等人(2022)采集了 1015 名 7~13 岁的儿童(454 名阅读障碍儿童)的汉字字符数据集, 采用多种算法对汉字的反应特征(如笔画、字素、音调等)、字符结构、回答特征(如正字法、语音词根等)、个人特征等分类变量进行了机器学习, 最终基于汉字字符的结构、书写正确率、词汇地位、笔画、音调、年级等核心特征构建了中文阅读障碍筛查模型。

值得注意的是, 眼动特征已经成为基于机器学习进行阅读障碍分类的常用指标, 它与机器学习的结合提供了认知过程的细粒度信息(Raatikainen et al., 2021), 可作为阅读障碍的高精度筛查工具。Bhargavi 和 Prabha (2020)收集了 185 名 9~10 岁儿童(97 名阅读障碍儿童)的眼动特征集用以建立阅读障碍的预测模型, 在此基础上采用多种机器学习算法提高预测精度, 发现具有较高准确率的最佳特征集是平均注视次数、平均注视时间、平均眼跳时间、总眼跳运动次数和平均注视次数。Ileri 等人(2022)记录了 33 名 9~10 岁儿童(20 名阅读障碍儿童)在阅读文本时的眼电图(electrooculography, EOG)信号, 通过机器学习分析了不同类型的眼球运动规律, 以此来筛查阅读障碍者。

随着智能移动设备的日益普及, 基于网络/手机游戏的数据收集技术拥有了广泛的用户基础。当前, 研究者已经开发了各种支持、检测和治疗阅读障碍的应用程序和游戏(Ahmad et al., 2022)。游戏化设计大多以语言能力、知觉加工、工作记忆、执行功能、阅读技能等为测量内容, 在形式上通过丰富的游戏元素来吸引和激励用户。Rello 等人(2020)设计了一款用于测查行为和认知缺陷的在线游戏来收集 3644 名 7~17 岁用户(其中包括 392 名阅读障碍患者)的数据, 从而建立了一个用于筛查阅读障碍的机器学习模型。Rauschenberger 等人(2022)通过网页游戏“MusVis”收集了 313 名儿童(7~12 岁, 其中包括 116 名阅读障碍儿童)玩游戏的节奏和频率, 在此基础上利用机器学习进行模型训练和预测。

阅读障碍的本质特征是大脑解剖结构中微妙的空间分布变化(Richlan et al., 2013; Tamboer et al., 2016; Vandermosten et al., 2012)。基于功能性磁共振成像(fMRI)、脑磁图(MEG)、脑电图(EEG)、正电子发射扫描(PET)等技术获取的大脑成像数据为阅读障碍的机器学习分类提供了客观证据(Da Silva et al., 2021; Ortiz et al., 2020; Thiede et al., 2020)。fMRI 的数据大多关注的是与语言和词汇决策相关的大脑区域, 探究个体在阅读任务期间大脑激活的功能差异(Chimeno et al., 2014)。Zahia 等人(2020)收集了 55 名 9 至 12 岁西班牙儿童(其中包括 18 名阅读障碍)的 fMRI 结构像, 使用深度学习算法对阅读障碍儿童进行自动识别。Da Silva 等人(2021)从巴西说葡萄牙语的 32 名 8~12 岁儿童(16 名阅读障碍儿童)中收集高分辨率的 T1-w 图像, 使用深度学习算法对视觉表征的重要区域进行分类。EEG 能够在保持时间和频域的情况下记录高时间分辨率的大脑信号, 反映儿童认知处理过程中的大脑功能状态, 为阅读障

表 1 机器学习在发展性阅读障碍儿童早期筛查中的应用

序号	国家	作者	被试			主要数据指标	算法/方法	评价指标(%) Accuracy/Specificity/ Precision/F1
			样本量	年龄	性别(男%)			
1	沙特阿拉伯	Ahmad et al. (2022)	3644 (392DD+non)	7~17 岁	50.8	语音能力, 认知能力	SVM, ANN	Accuracy = 95
2	沙特阿拉伯	AlGhamdi (2022)	3644 (392DD+non)	7~17 岁	50.8	语音能力, 认知能力	KNN, BT	Accuracy = 99.9
3	希腊	Asvestopoulou et al. (2019)	32DD, 37TD	8.5~12.5 岁	/	眼动特征	LSVM	Accuracy = 84.21
4	印度	Bhargavi & Prabha (2020)	97DD, 88non	9~10 岁	/	眼动特征	Hybrid SVM-PSO, LR, RF, KNN	96.6/95/95/96.6
5	荷兰	Chen et al. (2017)	495FR, 498TD	17~35 个月	54.6	词汇发展(N-CDI 量表)	SVM	68/70/67/-/-
6	中国	Cui et al. (2016)	28DD, 33TD	av11.6 岁;av11.8 岁	52.5	MRI (白质形态学特征)	LSVM	83.61/75/90.91/-/-
7	巴西	Da Silva et al. (2021)	16DYS, 16TYP	8~12 岁	62.5	fMRI (大脑激活差异)	2D CNN, 3D CNN, SVM	Accuracy = 94.83
8	西班牙	Formoso et al. (2021)	16DG, 32CG	88~100 个月	50	EEG (脑功能网络)	NB	90/93/86/-/-, AUC = 0.95
9	法国	Hmimdi et al. (2021)	46DA, 41non	av15.52 岁; av14.78 岁	56.3	眼动特征	LR, SVM	81.25/85/82.5/-/-
10	土耳其	Iler et al. (2022)	20DD, 13HC	8~11 岁	48.5	EOG (垂直/水平眼电信号)	1D-CNN	98.70/97.52/99.90/-/98.69
11	土耳其	Latifoglu et al. (2021)	20DD, 10HC	8~12 岁	50	EOG (眼电信号频谱图)	2D-CNN	99/100/98.2/-/98.9
12	中国香港	Lee et al. (2022)	454DD, 561TD	7~12 岁	/	汉字字符, 个人特征	NB, SVM, KNN, DT, ANN, LR	Accuracy = 78
13	波兰	Plonski et al. (2017)	130DG, 106CG	8.5~13.7 岁	53.0	MRI (灰质形态学特征)	LR	Accuracy = 65, AUC = 0.66
14	印度	Prabha & Bhargavi (2019)	97DD, 88non	9~10 岁	/	眼动特征	SVM-PSO, L SVM	95/100/89/-/-
15	澳大利亚	Radford et al. (2021)	43DD, 50non	7~14 岁	/	单词音频信号	LSTM, LR, SVM, KNN, RF, DT	Accuracy = 80
16	德国	Rauschenberger et al. (2022)	116DG, 197CG	7~12 岁	51.1	听觉特征, 视觉特征	RF, ET, GB	74/-/78/75
17	马来西亚	Rello et al. (2020)	3644 (392DD+non)	7~17 岁	50.8	语音能力, 认知能力	RF	Accuracy = 89.2
18	以色列	Shamir et al. (2019)	81RD, 44TR	6~14 岁	40	语音能力, 认知能力	SVM	-/75/75/-/-
19	德国	Skeide et al. (2016)	141 (37DD+non)	3~12 岁	59.6	MRI (白/灰质形态学特征), 基因	SVM	Accuracy = 75.53
20	伊朗	Tolami et al. (2021)	29DD, 25non	8~11 岁	/	语言特征	NB, MLP, KNN, SVM, LR, DT	93.33/-/-/94.44/93.21
21	塞尔维亚	Vajs et al. (2022)	15DD, 15non	7~13 岁	36.7	眼动特征	LR, SVM, KNN, RF	Accuracy = 94, AUC = 0.98
22	中国	Wang & Bi (2022)	187DD, 212TD	7~13 岁	64.2	与阅读相关的认知能力, 个人特征	GA-BPNN	Accuracy = 94.1, AUC = 0.95
23	美国	Yu et al. (2022)	227SAC	8~14 岁	/	虚拟迷宫数据, 个人特征, 阅读水平	RUSBoosted Trees	Accuracy = 70
24	西班牙	Zahia et al. (2020)	19DXR, 3 8non (TD+ MVR)	9~12 岁	58.2	fMRI (大脑激活区域)	3D CNN	72.3/71.43/60/67
25	马来西亚	Zainuddin et al. (2019)	10PDS, 10CDS, 10CG	7~12 岁	/	EEG (beta/theta 频带信息)	KNN, ELM	Accuracy = 89

注: 国家以第一作者所在国家为准, 算法/方法采用重点研究算法, 评价指标采用最佳算法的最优参数。

TD/CD/TDR/TYP/TR/CG/HC/non: 典型发展儿童; DG/DD/DYS/DA/RD: 阅读障碍儿童; PDS (poor dyslexic subject): 差阅读障碍儿童; CDS (capable dyslexic subject): 有能力的阅读障碍儿童; FR (family risk): 有阅读障碍家庭风险儿童; SAC (school-aged children): 学龄儿童; MVR: 单目视觉儿童(无阅读障碍); av (Average): 平均年龄; N-CDI: 荷兰版的 McArthur-Bates 沟通发展量表; SVM (Support Vector Machines): 支持向量机; KNN (K-Nearest Neighbors): K 最近邻; LR (Logistic Regression): 逻辑回归; CNN (Convolutional Neural Network): 卷积神经网络; RF (Random Forest): 随机森林; ET (Extra Trees): 极限森林; NB (Naïve Bayes): 朴素贝叶斯; DT (Decision Tree): 决策树; ANN (Artificial Neural Network): 人工神经网络; BT (Boosted Trees): 提升树; GA-BPNN: 遗传算法−反向传播神经网络; SVM−PSO: 粒子群算法优化支持向量机; GB (Gradient Boosting): 梯度提升; ELM (Extreme Learning Machine): 极限学习机; LSTM (Long-Short Term Memory neural networks): 长短期记忆神经网络; MLP (Multilayer Perceptron Neural Network Model): 多层感知机。

碍的早期诊断提供有效特征。研究者大多关注EEG信号的5个波段, 即delta, theta, alpha, beta和gamma(Ortiz et al., 2020), 通过脑电图通道之间的相位同步探究大脑的连通性, 然后提取鉴别特征用于阅读障碍的识别。Zainuddin等人(2019)采集了7~12岁的10名中度阅读障碍儿童、10名重度阅读障碍儿童和10名对照组儿童的EEG信号, 通过写作任务以K最邻近(KNN)和极限学习机(ELM)来筛查阅读障碍。Formoso等人(2021)采集了7~8岁的48名儿童(16名阅读障碍)的EEG信号, 测量通道之间的相位同步, 以揭示听觉处理过程中激活的脑功能网络。然后, 使用矢量量化无监督学习和贝叶斯算法相结合的方法提取鉴别特征, 用于阅读障碍的鉴别。在中文阅读障碍研究中, Cui等人(2016)采用结构磁共振成像(MRI)和扩散张量成像(DTI)收集了61名10~14岁学龄儿童(其中28名阅读障碍儿童)的3D T1-w图像(MPRAGE), 使用机器学习算法将阅读障碍儿童与典型发展儿童区分开来。

如今, 越来越多的研究者开始不局限于某种单一模态的数据收集, 他们将量表、行为、影像等数据进行整合, 试图提高阅读障碍及其生物标志物检测的准确性。纳入分析的25篇文献中使用的数据类型占比如下: 标准化心理教育测试和眼动特征数据各为28%, 其次是游戏测试数据为16%, MRI数据为12%, fMRI数据和EEG数据各为8%。

3.2　数据预处理、特征提取和特征选择

数据预处理的主要目的是使算法能够从数据集中提取最相关的可解释特征(Usman et al., 2021)。对于传统的机器学习方法, 预处理的第一步是将数据转换为定量(数字)或定性(文本类别)格式。也有一些量表或行为数据采用手动预处理方式, 如请专家将数据标记为无阅读障碍组和阅读障碍组(Khan et al., 2018)。在脑成像研究中, 研究者直接收集的数据通常是高维度多变量的数据。以64个通道的EEG数据为例, 即使在一个通道上计算得到一个指标, 则至少得到64个特征值。当特征值数量大于样本数量, 使用机器学习容易造成过拟合问题以及降低训练和预测速率。因此, 需要将高维度的特征降低到低维度的特征, 加快后续机器学习的分类和训练。例如, EEG信号预处理中常采用主成分分析(Principal Components Analysis, PCA)剔除数据次要成分的维度, 做到数据的降维(Ahire et al., 2022)。此外, 脑成像数据还可以使用不同类型的软件工具包进行预处理。如fMRI图像可使用matlab的SPM工具箱自动分割出不同的组织类型, 提高数据预处理时像素和体素的可比性(Zahia et al., 2020); 或者使用FreeSurfer图像分析套件提取可靠的皮层体积和厚度(Plonski et al., 2017)。

预处理完成的下一步是特征选择和提取, 目的是从原始特征中生成最相关、信息量最大的特征(Abd Rahman et al., 2020), 形成分类所需的数据集。标准化心理教育测试可选择的特征一般有问卷/认知测试分数、书写数据、语音数据等。眼动数据通常使用统计度量、基于离散和基于速度的算法选择特征, 使用主成分分析法(Principal Component Analysis, PCA)提取特征。fMRI数据的特征提取是从脑组织属性中提取大脑皮层属性特征, 常见的特征有: 皮层厚度、体积信息、各向异性分数和激活模式等。在EEG数据中, 一般使用傅里叶变换和小波变换提取信号的时间和频率信息。另外, 最近出现了一些新的特征提取方法, 如深度学习通过构造不同的网络结构自动从数据中进行特征提取, 具有良好的稳健性和较强的高维数据处理能力。例如, 在Ileri等人(2022)的研究中, 卷积神经网络(CNN)提供了输入的分段EOG信号的自动分类, 无需手动提取特征。表2总结了纳入分析的25篇文献中的特征类型。

3.3　模型训练与分类

在特征提取与选择完成之后, 研究者便可以利用机器学习进行模型训练与分类。机器学习大致分为两种类型: 无监督学习和有监督学习。无监督学习用于在不使用任何输出数据的情况下查找输入数据中的模式, 而监督学习主要用于预测未来事件(Russell & Norvig, 2010)。在监督学习中, 训练模型的目的是从标记的数据学习所有权重和偏差的理想值。近年来的研究一般使用监督学习算法探究阅读障碍患者和典型发展人群的分类问题。常见的算法有: 支持向量机(Support Vector Machines, SVM)、决策树(Decision Tree, DT)、随机森林(Random Forest, RF)、线性回归(Linear Regression, Linear-R)、逻辑回归(Logistic Regression, LR)、线性判别分析(Linear Discriminant Analysis, -LDA)、朴素贝叶斯(Naïve Bayes)、K最近邻(K-Nearest

表 2 机器学习在发展性阅读障碍儿童早期筛查中的特征类型

一级维度	具体特征
标准化心理教育测试	语音加工：正字法意识，语素意识，音节意识，字母意识，快速自动命名，语音意识，语音感知，语音流畅性，语言表达与理解，语法错误，语法复杂性，语法可读性，词汇多样性； 阅读能力：阅读速度，阅读时间，阅读流畅性，阅读准确性，阅读理解性； 认知能力：工作记忆，视觉/听觉工作记忆，视觉/听觉辨别和分类，视觉注意，加工速度； 字符特征：字符结构，书写正确率，词汇地位，部首，正字法，笔画，语音，首音，韵律，声调； 个人特征：年级，年龄，性别，智力； 其他特征：沟通发展量表(N-CDI)数据，虚拟迷宫学习数据；
眼动追踪	常见的眼动特征：注视时间/次数/频率/中位数长度/交叉变异性/分形维数，眼跳时间/次数/中位数长度/变异性/距离，眨眼次数/频率，回视，落点位置，双眼位置； 其他眼动特征：主动阅读时间，阅读速度/时间/错误率，垂直/水平 EOG 信号，EOG 信号的频谱图；
网络/手机游戏	语音能力：语言技能，语言练习； 认知能力：工作记忆，执行功能，感知过程； 听觉特征：持续时间，平均点击时间，间隔时间，命中率； 视觉特征：点击次数，点击时间
脑成像技术	MRI：白质的体积/各向异性分数/平均弥散率/轴向扩散系数/径向扩散系数，灰质的体积/厚度/面积/折叠指数/平均曲率； fMRI：阅读任务下的大脑激活区域，大脑激活差异； EEG：脑功能网络、beta/theta 频带信息
其他	阅读障碍的候选基因

Neighbors, KNN)、人工神经网络(Artificial Neural Network, ANN)和卷积神经网络(Convolutional Neural Network, CNN)等。在训练模型前，通常要将整个数据集分为测试集(testing set)和训练集(training set)。现有关于阅读障碍的机器学习的研究大多使用 K 折交叉验证(K-fold cross-validation)，将数据集分成 K 等份，其中 K−1 份用于训练集，1 份用于测试集，以 K 次测试结果的平均值作为最终的性能评估结果。例如，Plonski 等人(2017)采用 10 倍交叉验证法，AlGhamdi (2022)采用 5 倍交叉验证法和 20 倍交叉验证法。当样本量较小时，一些研究者也会选用 K 折交叉验证的特殊形式——留一法(Leave-one-out cross-validation)构建模型和评估分类结果(Cui et al., 2016; Asvestopoulou et al., 2019).

阅读障碍的识别问题在本质上是二元分类问题，即区分用户是否为阅读障碍者。SVM 的原理是从解决线性二分类问题出发，可以为超高维且特征空间稀疏的数据提供良好的性能。因此，SVM 成为阅读障碍研究中应用得最为广泛的算法。Shamir 等人(2019)采用 SVM 算法对基于标准化测试和 zippy 6 筛选测试的阅读评估数据进行分类，获得了 75%的特异性和灵敏度。Prabha 和 Bhargavi (2019)提出一种粒子群算法优化支持向量机(SVM-PSO)模型用于从眼动特征中提取阅读障碍的生物标记物。与线性支持向量机(Linear SVM)模型相比，该模型的预测准确率达到了 95%。此外，研究者将 SVM 混合其他算法来识别阅读障碍儿童。例如，使用 RF 选择最重要的特征作为 SVM 的输入，模型达到了 89.7%的准确率和 84.8%的召回率(Raatikainen et al., 2021)。

对于大数据间的复杂规律的挖掘来说，深度学习的出现解决了这一难题。深度学习算法具有更多的层次结果，因此对事物的建模或抽象表现能力更强，也能模拟更复杂的模型。ANN 是深度学习的基础，它模拟了大脑神经网络结构和功能，在不确定的识别(如语音识别、图像识别)中尤其有效(Lucchiari et al., 2014)。Ahmad 等人(2022) 使用 ANN 模型对综合游戏数据进行分类，获得了 95%的检测准确率。随着神经网络的发展，深度学习从浅层的 ANN 中衍生出来。其中，CNN 是用于阅读障碍分类的最受欢迎的深度学习模型(Usman

et al., 2021)。Da Silva 等人(2021)选取两种网络可视化技术在 CNN 输入层中学习高级特征, 仅从脑成像数据(fMRI)就对阅读障碍儿童的大脑状态进行了精准分类, 达到 94.8%的准确率。不仅如此, 研究者提出一种新的基于 EOG 信号的 CNN 方法来识别阅读障碍。Latifoglu 等人(2021)通过阅读时的跳线和返回眼球运动来筛查和跟踪阅读障碍儿童。他们使用二维卷积神经网络(2D-CNN)模型对这些频谱图图像进行分类, 获得了 99%的准确率、100%的灵敏度、98.18%的特异性和 98.94%的 F 分数。Ileri 等人(2022) 从水平和垂直通道记录 EOG 信号, 应用一维卷积神经网络(1D-CNN)对这两个通道的信号进行分类, 准确率分别为 98.70%和 80.94%。

事实上, 没有任何一种算法能够成为适用于所有数据集的最佳算法。算法的选择受到问题性质、数据集特征和数量、数据格式、训练和预测时间、存储需求等多种因素的影响。因此, 研究者越来越倾向于在多种算法中寻找最佳算法以获取最优参数。研究的整体趋势为从单一的传统机器学习算法走向深度学习算法(Deep neural network, DNN)以及比较多种不同类型的算法。Tolami 等人(2021) 以语言特征为分类指标, 构建了 NB, KNN, SVM, LR, DT 和 MLP 模型, 其中属于深度学习的 MLP 算法的最高分类准确率达到 93.33%。在中文阅读障碍的研究中, Lee 等人(2022)以汉字字符和个人特征为分类指标, 运用 NB、KNN、SVM、DT、LR 和 ANN 分别构建模型, 发现这 6 种算法都可以将阅读障碍儿童与典型发展儿童区分开来。其中, SVM 获得了 80.0%的最高准确率。

基于纳入分析的 25 篇文献, 算法的使用频次占比统计如下: SVM 占比为 27.3%, 其次是 KNN 和 LR 各占 12.7%, CNN 和 RF 各占 9.1%, NB 和 DT 各占 5.5%, ANN 占比为 3.6%, BT、GA-BPNN、GB、ELM、LSTM、MLP、RUSBoosted 和 ET 各占 1.8%。

3.4 性能评估

阅读障碍的结局变量均为二元分类变量, 对于二分类结果的评估首先需要对于不同样本分类的分类结果进行 4 类划分: 真阳性(True Positive, TP)、真阴性(True Negative, TN)、假阳性(False Positive, FP)、假阴性(False Negative, FN)。接着, 根据数据的类别划分定义评估指标。对于二分类问题, 最常用的评估指标是整体准确率。但是, 准确率只能体现正(阳性)、负(阴性)类样本合计的正确识别数占总样本数的比例。在实际应用中, 尤其是临床筛查中往往会出现数据中的正负样本量比例过大的问题。针对这些不平衡数据的分类问题, 需要采用多个指标对分类模型进行性能评估。其他常用的评估指标有: 灵敏度(Sensitivity)、特异性(Specificity)、精度(Precision)召回率(Recall)、F1 分数(F1 score)、Kappa 系数、AUROC 曲线与 P-R 曲线、阳性预测值(Positive Predictive Value, PV+)与阴性预测值(Negative Predictive Value, PV−)等。基于纳入分析的 25 篇文献, 机器学习模型的评估性能(以准确率为参考)总结如下: 标准化心理教育测试在 68%~94.1%之间; 眼动追踪测试在 81.25%~99%之间; 游戏测试在 74%~99.9%之间; 基于 EEG 捕获的脑成像数据在 89%~90%之间; 基于 fMRI 捕获的脑成像数据在 65%~94.83%之间。

4 基于机器学习的阅读障碍早期筛查的研究应用

4.1 发现阅读障碍的预测因素

机器学习最主要的优势在于模型的灵活度, 即可以拟合相当复杂的多项交互关系或者非线性关系, 由此产生令人瞩目的预测准确性。特别是研究涉及到预测性问题, 如预测微博用户的自杀风险、抑郁症的易感人群等, 机器学习的统计效果尤其突出。

基于机器学习有助于发现阅读障碍的预测因素, 我们可以有效检出具有阅读障碍风险的儿童对其进行及时干预, 从而减小儿童识字后甚至成年后阅读失败的可能性。例如, Tamboer 等人(2016)借助 MRI 技术构建 SVM 预测模型, 发现最可靠的分类体素位于左侧枕叶梭状回(Left Occipital Fusiform Gyrus, LOFG)、右侧枕叶梭状回(Right Occipital Fusiform Gyrus, ROFG)和左侧顶下小叶(Left Inferior Parietal Lobule, LIPL), 敏感性达到 82%, 特异性达到 78%。因此, 这些脑区是与阅读、拼写和语音相关的阅读障碍类型的潜在生物标志物。Prabha 和 Bhargavi (2019)基于 SVM-PSO 构建的预测模型显示, 平均注视次数、平均注视时间、平均扫视时间、总扫视次数和平均注视次

数等眼动特征可以作为儿童阅读障碍的风险预测指标。该模型预测的准确率高达 95%。Formoso 等人(2021)收集儿童的 EEG 信号，通过脑电通道之间的相位同步来表示脑区之间的连通性。研究结果显示在 16 Hz 的刺激下，alpha 和 beta 波段的辨别能力最强，AUC 值达到 0.95。在中文阅读障碍的研究中，Wang 和 Bi (2022)构建了基于遗传算法−反向传播神经网络模型(GA-BPNN)的中文阅读障碍预测模型，发现阅读准确性是预测汉语阅读困难儿童的最强因素，语音意识、假字准确率、语素意识、阅读流畅性、快速数字命名和非字符反应时间对预测也具有重要贡献。基于纳入本次系统综述的文献，最具预测性的特征总结详见表 3。

4.2 辅助筛查阅读障碍儿童

阅读障碍儿童的传统筛查主要通过专业的医疗机构和科研机构进行，方法以标准化心理教育测试结合儿童外在行为和家长报告为主。虽然近年来眼动追踪和脑成像技术为阅读障碍筛查提供了更加客观的技术支持，但通过这些复杂的测量工具来对每一个阅读障碍患者进行大范围识别几乎是不可能的。同时，这些测量工具还存在着价格昂贵、耗时长、普及性差、就诊渠道窄等弊端。为此，机器学习被用来辅助临床筛查和自动化识别，不仅可以纳入大量客观分类指标提高准确率，而且方便快捷，降低等待成本。

Asvestopoulou 等人(2019) 开发了一款阅读障碍的筛查工具 DysLexML，通过眼动追踪记录儿童默读期间的注视点，应用 LSVM 构建筛查模型，准确率达到 97%。值得一提的是，DysLexML 在存在噪声的情况下依然具有良好的稳健性和准确率。因此，DysLexML 可以覆盖更多数量和更多样化的人群，为在控制较少、规模较大的环境中(如幼儿园)开发廉价的眼动筛查工具奠定了基础。

以往研究认为，阅读障碍与虚拟 Hebb-Williams 迷宫任务的表现相关，但是通过实时观察任务表现来对阅读障碍儿童进行分类尚不可行(Gabel et al., 2021)。Yu 等人(2022)基于机器学习算法开创性地根据迷宫任务的表现预测阅读能力，实现以阅读风险的百分比形式实时反馈阅读成绩。他们以虚拟迷宫数据、阅读水平和个人特征为指标，利用 RUSBoosted 树算法(RUSBoosted Trees algorithm)构建的模型分类准确率达到 70%以上。随着计算机网络的发展和电子产品的普及，以在线评估平台或应用程序初步筛查阅读障碍的方式逐渐流行起来。研究者开发了一款分析真实环境下音频信号的应用程序 Dystech，他们发现与适当的音频信号处理相关的机器学习算法可以提取出人类专家无法获取的模式，筛查准确率达到 80%(Radford et al., 2021)。Rauschenberger 等人(2022)设计了一款语言独立游戏 MusVis，使用 RF 和额外树(Extra trees, ET)对收集到的游戏的数据进行训练，德语和西班牙语阅读障碍的分类准确率分别达到 74%和 69%。

4.3 预测阅读障碍高风险儿童

人们普遍认为阅读障碍具有遗传基础(约 70%) (苏萌萌 等, 2012; Galaburda et al., 2006)。即使部分有家庭风险的儿童没有出现阅读障碍，但是他们在拼写、非单词阅读和阅读理解等任务上的表现仍然比典型发展儿童要差(Lyytinen et al., 2005)。在早期发现具有高风险的阅读障碍的儿童将使早期预防和干预成为可能。这种早期预测功能可通过训练机器学习预测模型来实现。Skeide 等人(2016)认为与识字相关的重要脑区的神经可

表 3 机器学习在发展性阅读障碍儿童早期筛查中最具预测性的特征

一级维度	最具预测性的特征
标准化心理教育测试	阅读准确性，笔画，词汇地位，字符结构，词汇量，年级;
眼动追踪	异向眼动，注视交叉变异性，注视次数/持续时间，眼跳次数/持续时间/长度，短前向移动个数，多重注视词个数，水平 EOG 信号;
网络/手机游戏	视觉/听觉点击总次数，视觉/听觉第一次点击时间;
脑成像技术	MRI: NRSN1 相关的“视觉单词形成区”的灰质体积，假定阅读系统内/边缘系统/运动系统的白质区域，左半球外侧裂周区有额外褶皱的非典型曲率模式; fMRI: 左侧枕叶，顶下小叶，右侧前额叶; EEG: beta 频带信息，db2 小波，db4 小波

塑性可能受到遗传变异的调控，从而预先限制了儿童的读写能力。为此，他们采集了 4~8 年级儿童和幼儿园~1 年级儿童的灰/白质体积以及与识字相关的基因信息，利用 LSVM 构建了阅读障碍的预测模型，其准确率分别达到了 73%和 75%。Chen 等人(2017) 根据词汇发展量表分析了 17~35 个月的儿童的词汇总量和词汇类别的群体水平差异，使用SVM来对家庭风险的阅读障碍儿童和典型发展儿童进行分类。研究结果显示风险预测模型的准确率为 68%，敏感性为 70%，特异性为 67%，表明通过机器学习可以在识字前的早期阶段区分出有家庭风险的阅读障碍儿童和典型发展儿童。这个阶段有家庭风险的阅读障碍儿童的颞顶叶和颞枕叶区域已经显示出功能和结构上的改变，并且类似于阅读障碍患者中观察到的变化(Hosseini et al., 2013; Kraft et al., 2015)。

5　机器学习应用于儿童阅读障碍早期筛查的优势与不足

5.1　优势

近年来，机器学习在阅读障碍及其生物标志物检测中的应用越来越受到研究者的青睐，其优势主要体现在以下三个方面。第一，机器学习可以识别变量之间复杂的非线性关系，提供对阅读障碍更加精准的筛查与发展性预测。阅读障碍是多种因素相互作用的结果(Morris et al., 1998)，传统的统计学方法(如逻辑回归)所确定的单个或多个预测因子存在预测能力弱或是无法体现因子间的交互作用的缺点，不能对数据进行充分挖掘。机器学习则更适合分析阅读障碍这一类结构复杂问题。以反向传播神经网络(BPNN)为例，BPNN 作为人类大脑工作机制的模拟，不仅可以处理模糊映射关系，还可以识别变量之间复杂的非线性关系(Lyu & Zhang, 2019)。无论是在字母语言还是汉语中，通过采集与阅读相关的认知测试或语音测试数据，BPNN 模型皆可有效筛查出阅读障碍儿童(Wang & Bi, 2022)。第二，与人为识别阅读障碍的方式相比，机器学习一方面避免了主观理解偏差的影响，另一方面能够自动化重复的任务，在更短的时间内分析更多的数据，实现比人工算法更高的准确性和可重复性。第三，机器学习具备强大的高维数据处理能力，可从脑成像数据中提取出额外的、关键性的区辨性信息，检测到人眼无法观测到的可能反映重要病理生理机制的微小成像的异常。大脑功能和发育的差异是阅读障碍风险的早期迹象。随着年龄增长，突触的快速形成使得儿童大脑的激活模式发生变化，但除非严重受伤或病危，大脑结构从童年到成年保持不变。因此，高维的脑成像数据可为阅读障碍的识别提供更准确的结果。例如，Da Silva 等人(2021) 从大脑成像对发展性阅读障碍儿童进行了 94.8%的准确分类，同时利用特征可视化技术(CAM)和基于梯度的特征可视化技术(Grad-CAM)在卷积神经网络层负责学习高级特征，提供了阅读障碍儿童和典型发展儿童在阅读的策略控制和注意过程中相关的大脑区域的可视化图像。这种在切片水平上对大脑状态的预测，以及随后生成的与分类相关的更细粒度的特征信息可以提高模型的可解释性。

5.2　不足和启示

首先，缺少对最佳干预期的被试群体研究。阅读障碍具有可遗传性，68%的同卵双胞胎和高达 40%~60%的一级亲属之间共同患有阅读障碍(Vogler et al., 1985)。几个阅读障碍的候选基因，如 ROBO1，DCDC2，DYX1C1，KIAA0319，已经证实在儿童的大脑发育中发挥着重要作用(Galaburda et al., 2006)。儿童的大脑可塑性增强的早期阶段处于幼儿园至小学 1 年级期间，是阅读障碍早期干预的最佳时期(Fox et al., 2010)。研究发现，对幼儿园和 1 年级的高危阅读障碍儿童进行有效干预的效果(平均效应量为 0.31~0.84)远高于 2 年级和 3 年级的高危阅读障碍儿童(平均效应量为 0.23~0.27) (Wanzek & Vaughn, 2007)。 为此，在最佳干预期之前对阅读障碍儿童，尤其是对有家庭风险的阅读障碍儿童进行精准的早期识别至关重要。遗憾的是，基于上述回顾的所有文献，仅有一项机器学习的阅读障碍研究的被试年龄较小(17~35 个月)，其余研究中被试年龄段在 6~17 岁之间，儿童识字前(3~7 岁)这一阶段的研究几乎是空白[2]。未来研究需要在儿童识字前广泛收集他们及一级亲属的相关数据，关注遗传和环境中可能

[2] 笔者注：Skeide 等人(2016)的研究共 141 名被试，其中 20 名被试年龄为 5~6 岁。这 20 名被试在 5~6 之间进行 MRI 扫描，约 1.7 年后的 7~9 岁左右进行识字能力评估，因此在此处未纳入 3~7 岁的范畴。

的风险因素，建立多模态数据库，借助机器学习的强大分类功能筛选阅读障碍儿童并确立较为稳定的行为/生物标记物，最终搭建方便、快捷、精准、科学的早期筛查系统。

其次，机器学习研究中采集的数据质量参差不齐，采集标准不统一，数据样本不足。基于机器学习的阅读障碍数据库采集呈现单一数据库到多方数据库、单一模态到多模态的趋势。由于数据库来自不同的实验室和不同的人群，采集标准尚未统一，数据分布的特征不同，大量的数据不兼容、结构复杂。因此有必要建立标准化异构数据库，提高模型所需的计算力，避免资源浪费。采集标准不统一的现象尤其充斥于阅读障碍儿童的脑成像数据库。一方面，成像设备型号、参数不统一会对数据质量产生一定影响。由于缺乏权威、固定的标准，脑成像的可重复性得不到一致认可。另一方面，分类的准确率在很大程度上取决于样本量大小。相较于问卷、行为数据，各课题组公开/非公开的阅读障碍相关数据库中脑成像数据量较少。用小样本训练的模型很容易陷入对小样本的过拟合以及对目标任务的欠拟合。针对以上问题，首先可以通过国际合作建立数据采集以及数据共享的统一标准的平台，实现脑成像数据的可重复性应用。其次，可以通过增多训练数据、缩小模型需要搜索的空间和优化搜索最优模型的过程等方式进行补救。

再者，暂时无法在临床实践中达到高转化力并得到更广泛的使用。虽然大量的研究发现大脑形态、眼球运动和正常听觉系统中检测到的变化可以作为阅读障碍识别的神经生物标记物，但传统门诊对于阅读障碍的筛查依旧以标准化心理教育测试(行为标记)为主。这是由于标准化心理教育测试具有测验内容有代表性、标准化程度高、信效度高和使用方便经济等优点。机器学习目前尚不具备临床转化的必备条件。首先，训练数据欠缺代表性。实验数据通常是在控制实验无关变量的前提下寻求对典型样本的估计，但如果我们的目的是是创建可推广的预测算法，样本需囊括实际生活中大量的个体化病例。其次，机器学习模型可解释性和透明度低。存在“算法黑箱”，输入的数据和输出答案之间存在不可观察的空间，甚至开发人员都不能完全理解算法运作的具体细节。再者，机器学习的性能指标不具备临床适用性，如F1分数、召回率可能无法适用于临床环境，很难被临床医生和研究人员理解。最后，干预方法的验证研究不足。阅读障碍的早期筛查的最终目的是为了给儿童提供行之有效的早期干预。但是，目前仅有两项研究将机器学习与阅读障碍干预联系起来(Atkar & Jayaraju, 2021; Oliaee et al., 2022)。之前对基于机器学习的阅读障碍的EEG研究主要是通过脑电图的组间差异(特别是单个频段的功率)来区分阅读障碍儿童和典型发展儿童。Oliaee等人(2022)开创性地对特定治疗计划前后的阅读障碍儿童进行分类，为评估阅读障碍治疗方案的有效性提供了新的方法。他们利用PCA和序列前向选择(Sequential Floating Forward Selection, SFFS)算法，从记录的脑电图信号中提取出最优特征子集，发现阅读障碍儿童在接受经颅直流电刺激(Transcranial Direct Current Stimulation, tDCS)治疗和认知训练前后的脑电信号在不同区域的频谱和相位相关特征上发生了显著变化，最具识别力的特征子集的分类准确率达到92%。Atkar和Jayaraju (2021)使用一种深度学习−无监督学习的生成对抗网络模型(Generative Adversarial Networks, GAN)生成两个或三个字母的印地语单词的原始音频数据，使用生成的数据建立MelGAN系统。该系统通过让阅读障碍儿童重复单词的正确发音来加快恢复过程，旨在为教师提供一个有效的辅助工具。虽然使用机器学习评估干预效果以及辅助创建干预工具开始走进研究人员的视野，但它们的实用性和可验证性仍有待进一步提高。

最后，被试的数据安全和隐私保护受到威胁。机器学习模型训练需要大量数据，但数据库往往包含大量隐私数据，如个人身份信息、家庭信息等。如何低成本且高效地防止隐私泄漏变得极为重要。Usman和Muniyandi (2020)构建了一种基于CNN模型和余数模型(RNS)进行阅读障碍安全分类的方法。他们利用RNS的特殊模块集开发了一个像素−比特流编码器，在使用级联CNN进行分类之前对训练集和测试集中MRI数据的每个像素的7位二进制值进行加密，再使用加密测试数据集预测阅读障碍儿童。此外，在数据共享之前制定知情同意、伦理审核同样有利于防止潜在的数据滥用。

综上所述，机器学习已被逐渐应用于阅读障

碍的早期筛查中。数据采集方式从单一模态向多模态的异构数据转变, 并使用多种模型验证最佳分类效果, 分类性能在 67%~100%之间。当前使用最多的机器学习算法是 SVM, 未来深度学习有望为阅读障碍实现更高的分类性能。在应用中, 阅读障碍的机器学习研究仍存在样本量少、临床实践率低、多模态数据结合不足、分类性能有待提高等问题。并且, 缺少对最佳干预期的儿童群体研究, 没有真正实现阅读障碍儿童的早期筛查。未来的研究首先应重点关注学龄前儿童的风险识别, 着眼于阅读障碍的早期筛查的标记物研究。其次, 由于阅读障碍并不特定于某个地区、语言和文化, 因此需要开发独立于语言的数据收集方法以建立统一标准的阅读障碍数据库。最后, 未来的研究需要采集多个来源数据(如量表、行为、脑成像等)、混合多种模型以及考虑多模态的深度学习框架提高机器学习的预测力, 不断优化构建的阅读障碍筛查模型, 最终实现临床实践中的广泛使用。

参考文献

*为纳入系统分析的文献

苏萌萌, 张玉平, 史冰洁, 舒华. (2012). 发展性阅读障碍的遗传关联分析. *心理科学进展*, *20*(8), 1259−1267.

Aaron, P. G., Joshi, M., & Williams, K. A. (1999). Not all reading disabilities are alike. *Journal of Learning Disabilities*, *32*(2), 120−137. https://doi.org/10.1177/002221949903200203

Abd Rahman, R., Omar, K., Noah, S. A. M., Danuri, M. S. N. M., & Al-Garadi, M. A. (2020). Application of machine learning methods in mental health detection: A systematic review. *IEEE Access*, *8*, 183952−183964. https://doi.org/10.1109/ACCESS. 2020.3029154

Ahire, N., Awale, R. N., Patnaik, S., & Wagh, A. (2022). A comprehensive review of machine learning approaches for dyslexia diagnosis. *Multimedia Tools and Applications*, *82*, 13557−13577. https://doi.org/10.1007/s11042-022-13939-0

*Ahmad, N., Rehman, M. B., El Hassan, H. M., Ahmad, I., & Rashid, M. (2022). An efficient machine learning-based feature optimization model for the detection of dyslexia. *Computational Intelligence and Neuroscience*, *2022*, 8491753. https://doi.org/10.1155/2022/8491753

*AlGhamdi, A. S. (2022). Novel ensemble model recommendation approach for the detection of dyslexia. *Children-Basel*, *9*, 1337. https://doi.org/10.3390/children9091337

*Asvestopoulou, T., Manousaki, V., Psistakis, A., Smyrnakis, I., Andreadakis, V., Aslanides, I. M., & Papadopouli, M. (2019). DysLexML: Screening tool for dyslexia using machine learning. *arXiv*. https://doi.org/10.48550/arXiv. 1903.06274

Atkar, G., & Jayaraju, P. (2021). Speech synthesis using generative adversarial network for improving readability of Hindi words to recuperate from dyslexia. *Neural Computing & Applications*, *33*(15), 9353−9362. https://doi.org/10.1007/s00521-021-05695-3

Ballester, P. L., da Silva, L. T., Marcon, M., Esper, N. B., Frey, B. N., Buchweitz, A., & Meneguzzi, F. (2021). Predicting brain age at slice level: Convolutional neural networks and consequences for interpretability. *Frontiers in Psychiatry*, *12*, 598518. https://doi.org/10.3389/fpsyt.2021.598518

*Bhargavi, R., & Prabha, A. J. (2020). Predictive model for dyslexia from fixations and saccadic eye movement events. *Computer Methods and Programs in Biomedicine*, *195*(5), 105538. https://doi.org/10.1016/j.cmpb.2020.105538

Borleffs, E., Glatz, T. K., Daulay, D. A., Richardson, U., Zwarts, F., & Maassen, B. A. M. (2018). GraphoGame SI: The development of a technology-enhanced literacy learning tool for standard Indonesian. *European Journal of Psychology of Education*, *33*(4), 595−613. https://doi.org/10.1007/s10212-017-0354-9

Burns, M. K., VanDerHeyden, A. M., Duesenberg-Marshall, M. D., Romero, M. E., Stevens, M. A., Izumi, J. T., & McCollom, E. M. (2022). Decision accuracy of commonly used dyslexia screeners among students who are potentially at-risk for reading difficulties. *Learning Disability Quarterly*. Advance online publication. https://doi.org/10.1177/07319487221096684

Catts, H. W., McIlraith, A., Bridges, M. S., & Nielsen, D. C. (2017). Viewing a phonological deficit within a multifactorial model of dyslexia. *Reading and Writing*, *30*(3), 613−629. https://doi.org/10.1007/s11145-016-9692-2

Catts, H. W., & Petscher, Y. (2022). A cumulative risk and resilience model of dyslexia. *Journal of Learning Disabilities*, *55*(3), 171−184. https://doi.org/10.1177/00222194211037062

*Chen, A., Wijnen, F., Koster, C., & Schnack, H. (2017). Individualized early prediction of familial risk of dyslexia: A study of infant vocabulary development. *Frontiers in Psychology*, *8*, 156. https://doi.org/10.3389/fpsyg.2017.00156

Chimeno, Y. G., Zapirain, B. G., Prieto, I. S., & Fernandez-Ruanova, B. (2014). Automatic classification of dyslexic children by applying machine learning to fMRI images. *Bio-Medical Materials and Engineering*, *24*(6), 2995−3002. https://doi.org/10.3233/BME-141120

*Cui, Z. X., Xia, Z. C., Su, M. M., Shu, H., & Gong, G. L. (2016). Disrupted white matter connectivity underlying developmental dyslexia: A machine learning approach.

Human Brain Mapping, *37*(4), 1443−1458. https://doi.org/10.1002/hbm.23112

*Da Silva, L. T., Esper, N. B., Ruiz, D. D., Meneguzzi, F., & Buchweitz, A. (2021). Visual explanation for identification of the brain bases for developmental dyslexia on fMRI data. *Frontiers in Computational Neuroscience*, *15*, 584659. https://doi.org/10.3389/fncom.2021.594659

Farah, R., Ionta, S., & Horowitz-Kraus, T. (2021). Neuro-behavioral correlates of executive dysfunctions in dyslexia over development from childhood to adulthood. *Frontiers in Psychology*, *12*, 708863. https://doi.org/10.3389/fpsyg.2021.708863

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2019). *Learning disabilities: From identification to intervention, 2nd ed*. The Guilford Press.

*Formoso, M. A., Ortiz, A., Martinez-Murcia, F. J., Gallego, N., & Luque, J. L. (2021). Detecting phase-synchrony connectivity anomalies in EEG signals. Application to dyslexia diagnosis. *Sensors*, *21*(21), 7061. https://doi.org/10.3390/s21217061

Fox, S. E., Levitt, P., & Nelson, C. A. (2010). How the timing and quality of early experiences influence the development of brain architecture. *Child Development*, *81*(1), 28−40. https://doi.org/10.1111/j.1467-8624.2009.01380.x

Gabel, L. A., Voss, K., Johnson, E., Lindstrom, E. R., Truong, D. T., Murray, E. M., … Gruen, J. R. (2021). Identifying dyslexia: Link between maze learning and dyslexia susceptibility gene, DCDC2, in young children. *Developmental Neuroscience*, *43*(2), 116−133. https://doi.org/10.1159/000516667

Galaburda, A. M., LoTurco, J., Ramus, F., Fitch, R. H., & Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. *Nature Neuroscience*, *9*(10), 1213−1217. https://doi.org/10.1038/nn1772

Gilvary, C., Elkhader, J., Madhukar, N., Henchcliffe, C., Goncalves, M. D., & Elemento, O. (2020). A machine learning and network framework to discover new indications for small molecules. *PLoS Computational Biology*, *16*(8), e1008098. https://doi.org/10.1371/journal.pcbi.1008098

Hale, J., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., …Yalof, J. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly*, *33*(3), 223−236. https://doi.org/10.1177/073194871003300310

*Hmimdi, A., Ward, L. M., Palpanas, T., & Kapoula, Z. (2021). Predicting dyslexia and reading speed in adolescents from eye movements in reading and non-reading tasks: A machine learning approach. *Brain Sciences*, *11*(10), 1337. https://doi.org/10.3390/brainsci11101337

Hosseini, S. M. H., Black, J. M., Soriano, T., Bugescu, N., Martinez, R., Raman, M. M., … Hoeft, F. (2013). Topological properties of large-scale structural brain networks in children with familial risk for reading difficulties. *NeuroImage*, *71*, 260−274. https://doi.org/10.1016/j.neuroimage.2013.01.013

*Ileri, R., Latifoglu, F., & Demirci, E. (2022). A novel approach for detection of dyslexia using convolutional neural network with EOG signals. *Medical & Biological Engineering & Computing*, *60*(11), 3041−3055. https://doi.org/10.1007/s11517-022-02656-3

Kaisar, S. (2020). Developmental dyslexia detection using machine learning techniques: A survey. *ICT Express*, *6*(3), 181−184. https://doi.org/10.1016/j.icte.2020.05.006

Khan, R. U., Lee, J., & Yin, B. O. (2018). Machine learning and dyslexia: Diagnostic and classification system (DCS) for kids with learning disabilities. *Journal of Engineering Technology*, *7*(3), 97−100.

Kraft, I., Cafiero, R., Schaadt, G., Brauer, J., Neef, N. E., Mueller, B., … Skeide, M. A. (2015). Cortical differences in preliterate children at familiar risk of dyslexia are similar to those observed in dyslexic readers. *Brain*, *138*(9), e378. https://doi.org/10.1093/brain/awv036

Larco, A., Carrillo, J., Chicaiza, N., Yanez, C., & Luján-Mora, S. (2021). Moving beyond limitations: Designing the Helpdys App for children with dyslexia in rural areas. *Sustainability*, *13*, 7801. https://doi.org/10.3390/su13137081

*Latifoglu, F., Ileri, R., & Demirci, E. (2021). Assessment of dyslexic children with EOG signals: Determining retrieving words/re-reading and skipping lines using convolutional neural networks. *Chaos Solitons & Fractals*, *145*, 110721. https://doi.org/10.1016/j.chaos.2021.110721

*Lee, S. M. K., Liu, H. W., & Tong, S. X. (2022). Identifying chinese children with dyslexia using machine learning with character dictation. *Scientific Studies of Reading*. https://doi.org/10.1080/10888438.2022.2088373.

Livingston, E. M., Siegel, L. S., & Ribary, U. (2018). Developmental dyslexia: Emotional impact and consequences. *Australian Journal of Learning Difficulties*, *23*(2), 107−135. https://doi.org/10.1080/19404158.2018.1479975

Lucchiari, C., Folgieri, R., & Pravettoni, G. (2014). Fuzzy cognitive maps: A tool to improve diagnostic decisions. *Diagnosis (Berl)*, *1*(4), 289−293. https://doi.org/10.1515/dx-2014-0026

Lyu, J., & Zhang, J. (2019). BP neural network prediction model for suicide attempt among Chinese rural residents. *Journal of Affective Disorders*, *246*, 465−473. https://doi.org/10.1016/j.jad.2018.12.111

Lyytinen, P., Eklund, K., & Lyytinen, H. (2005). Language development and literacy skills in late-talking toddlers with and without familial risk for dyslexia. *Annals of Dyslexia*, *55*(2), 166−192. https://doi.org/10.1007/s11881-

005-0010-y

McGrath, L. M., Peterson, R. L., & Pennington, B. F. (2020). The multiple deficit model: Progress, problems, and prospects. *Scientific Studies of Reading*, *24*(1), 7−13. https://doi.org/10.1080/10888438.2019.1706180

Miciak, J., & Fletcher, J. M. (2020). The critical role of instructional response for identifying dyslexia and other learning disabilities. *Journal of Learning Disabilities*, *53*(5), 343−353. https://doi.org/10.1177/0022219420906801

Miciak, J., Stuebing, K. K., Vaughn, S., Roberts, G., Barth, A. E., & Fletcher, J. M. (2014). Cognitive attributes of adequate and inadequate responders to reading intervention in middle school. *School Psychology Review*, *43*(4), 407−427. https://doi.org/10.17105/SPR-13-0052.1

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., … Shaywitz, B. A. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology*, *90*(3), 347−373. https://doi.org/10.1037/0022-0663.90.3.347

Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., … Lyytinen, H. (2015). GraphoGame — A catalyst for multi-level promotion of literacy in diverse contexts. *Frontiers in Psychology*, *6*, 671. https://doi.org/10.3389/fpsyg.2015.00671

Oliaee, A., Mohebbi, M., Shirani, S., & Rostami, R. (2022). Extraction of discriminative features from EEG signals of dyslexic children; before and after the treatment. *Cognitive Neurodynamics*, *16*(6), 1249−1259. https://doi.org/10.1007/s11571-022-09794-2.

Ortiz, A., Martinez-Murcia, F. J., Luque, J. L., Gimenez, A., Morales-Ortega, R., & Ortega, J. (2020). Dyslexia diagnosis by EEG temporal and spectral descriptors: An anomaly detection approach. *International Journal of Neural Systems*, *30*(7). 2050029. https://doi.org/10.1142/S012906572050029X

*Plonski, P., Gradkowski, W., Altarelli, I., Monzalvo, K., van Ermingen-Marbach, M., Grande, M., … Jednorog, K. (2017). Multi-parameter machine learning approach to the neuroanatomical basis of developmental dyslexia. *Human Brain Mapping*, *38*(2), 900−908. https://doi.org/10.1002/hbm.23426

*Prabha, A. J., & Bhargavi, R. (2019). Prediction of dyslexia from eye movements using machine learning. *Iete Journal of Research, 68*(2), 814−823. https://doi.org/10.1080/03772063.2019.1622461

Raatikainen, P., Hautala, J., Loberg, O., Kärkkäinen, T., Leppänen, P., & Nieminen, P. (2021). Detection of developmental dyslexia with machine learning using eye movement data. *Array*, *12*, 100087. https://doi.org/10.1016/j.array.2021.100087

*Radford, J., Richard, G., Richard, H., & Serrurier, M. (2021, February). *Detecting dyslexia from audio records: An AI approach*. Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies−HEALTHINF (pp. 58−66), Electr Network. https://doi.org/10.5220/0010196000580066

*Rauschenberger, M., Baeza-Yates, R., & Rello, L. (2022). A universal screening tool for dyslexia by a web-game and machine learning. *Frontiers in Computer Science*, *3,* 628634. https://doi.org/10.3389/fcomp.2021.628634

*Rello, L., Baeza-Yates, R., Ali, A., Bigham, J. P., & Serra, M. (2020). Predicting risk of dyslexia with an online gamified test. *PLoS ONE*, *15*(12), e0241687. https://doi.org/10.1371/journal.pone.0241687

Richlan, F., Kronbichler, M., & Wimmer, H. (2013). Structural abnormalities in the dyslexic brain: A meta-analysis of voxel-based morphometry studies. *Human Brain Mapping*, *34*(11), 3055−3065. https://doi.org/10.1002/hbm.22127

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach*. Hoboken, NJ: Prentice Hall.

Sanfilippo, J., Ness, M., Petscher, Y., Rappaport, L., Zuckerman, B., & Gaab, N. (2020). Reintroducing dyslexia: Early identification and implications for pediatric practice. *Pediatrics*, *146*(1), e20193046. https://doi.org/10.1542/peds.2019-3046

*Shamir, N., Zivan, M., & Horowitz‐Kraus, T. (2019). Six‐minute screening test can provide valid information about the skills that underlie childhood reading and cognitive abilities. *Acta Paediatrica*, *108*(7), 1278−1284. https://doi.org/10.1111/apa.14680

Sihvonen, A. J., Virtala, P., Thiede, A., Laasonen, M., & Kujala, T. (2021). Structural white matter connectometry of reading and dyslexia. *NeuroImage*, *241,* 118411. https://doi.org/10.1016/j.neuroimage.2021.118411

*Skeide, M. A., Kraft, I., Mueller, B., Schaadt, G., Neef, N. E., Brauer, J., … Friederici, A. D. (2016). NRSN1 associated grey matter volume of the visual word form area reveals dyslexia before school. *Brain*, *139*, 2792−2803. https://doi.org/10.1093/brain/aww153

Tamboer, P., Vorst, H. C. M., Ghebreab, S., & Scholte, H. S. (2016). Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia. *Neuroimage-Clinical*, *11*, 508−514. https://doi.org/10.1016/j.nicl.2016.03.014

Thiede, A., Glerean, E., Kujala, T., & Parkkonen, L. (2020). Atypical MEG inter-subject correlation during listening to continuous natural speech in dyslexia. *NeuroImage*, *216*, 116799. https://doi.org/10.1016/j.neuroimage.2020.116799

*Tolami, F. A., Khorasani, M., Kahani, M., Yazdi, S. A. A., & Ghalenoei, M. A. (2021, October). *An intelligent linguistic error detection approach to automated diagnosis of Dyslexia disorder in Persian speaking children*. 11th International Conference on Computer Engineering and

Knowledge (ICCKE) (pp. 393−398), Mashad, Iran. https://doi.org/10.1109/ICCKE54056.2021.9721446

Usman, O. L., & Muniyandi, R. C. (2020). CryptoDL: Predicting dyslexia biomarkers from encrypted neuroimaging dataset using energy-efficient residue number system and deep convolutional neural network. *Symmetry-Basel*, *12*(5), 836. https://doi.org/10.3390/sym12050836

Usman, O. L., Muniyandi, R. C., Omar, K., & Mohamad, M. (2021). Advance machine learning methods for dyslexia biomarker detection: A review of implementation details and challenges. *IEEE Access*, *9*, 36879−36897. https://doi.org/10.1109/ACCESS.2021.3062709

*Vajs, I., Kovic, V., Papic, T., Savic, A. M., & Jankovic, M. M. (2022). Spatiotemporal eye-tracking feature set for improved recognition of dyslexic reading patterns in children. *Sensors*, *22*(13), 4900. https://doi.org/10.3390/s22134900

Vandermosten, M., Boets, B., Wouters, J., & Ghesquiere, P. (2012). A qualitative and quantitative review of diffusion tensor imaging studies in reading and dyslexia. *Neuroscience and Biobehavioral Reviews*, *36*(6), 1532−1552. https://doi.org/10.1016/j.neubiorev.2012.04.002

Vogler, G. P., Defries, J. C., & Decker, S. N. (1985). Family history as an indictor of risk for reading disability. *Journal of Learning Disabilities*, *17*(10), 616−618. https://doi.org/10.1177/002221948401701009

Walda, S., Hasselman, F., & Bosman, A. (2022). Identifying determinants of dyslexia: An ultimate attempt using machine learning. *Frontiers in Psychology*, *13*, 869352. https://doi.org/10.3389/fpsyg.2022.869352

*Wang, R., & Bi, H. Y. (2022). A predictive model for chinese children with developmental dyslexia − Based on a genetic algorithm optimized back-propagation neural network. *Expert Systems with Applications*, *187*, 115949. https://doi.org/10.1016/j.eswa.2021.115949

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*(4), 541−561. https://doi.org/10.1080/02796015.2007.12087917

Yang, X., Zhang, J., Lv, Y., Wang, F., Ding, G., Zhang, M., … Song, Y. (2021). Failure of resting-state frontal−occipital connectivity in linking visual perception with reading fluency in Chinese children with developmental dyslexia. *NeuroImage*, *233*, 117911. https://doi.org/10.1016/j.neuroimage.2021.117911

*Yu, Y. C., Shyntassov, K., Zewge, A., & Gabel, L. (2022, March). *Classification predictive modeling of dyslexia*. 56th Annual Conference on Information Sciences and Systems (pp. 177−181), Electr Network. https://doi.org/10.1109/CISS53076.2022.9751182

*Zahia, S., Garcia-Zapirain, B., Saralegui, I., & Fernandez-Ruanova, B. (2020). Dyslexia detection using 3D convolutional neural networks and functional magnetic resonance imaging. *Computer Methods and Programs in Biomedicine*, *197*, 105726. https://doi.org/10.1016/j.cmpb.2020.105726

*Zainuddin, A. Z. A., Mansor, W., Lee, K. Y., & Mahmoodin, Z. (2019, July). *Comparison of extreme learning machine and K-nearest neighbour performance in classifying EEG signal of normal, poor and capable dyslexic children*. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 4513−4516), Berlin, Germany. https://doi.org/10.1109/EMBC.2019.8857569

Application of machine learning in early screening of children with dyslexia

BU Xiaoou, WANG Yao, DU Yawen, WANG Pei

(*Department of Special Education, Faculty of Education, East China Normal University, Shanghai* 200062, *China*)

Abstract: Developmental dyslexia is the most prevalent form of specific learning disorder with a neurobiological basis that not only restricts an individual's academic achievement and career development, but also negatively affects an individual's psychological and social adjustment substantially. Recently, machine learning has been gradually applied to the early screening of children with dyslexia due to its powerful data processing and mining capabilities, accumulating richer results in various aspects such as standardized psychoeducational testing, eye tracking, game testing and brain imaging. However, machine learning still has limitations in terms of participant selection, data collection, transformation potential, security and privacy. Future researches need to focus on the early identification of pre-school children with dyslexia, construct multimodal data, and find the best classifier among multiple classifiers to obtain optimal parameters, which will eventually achieve widespread use in clinical practice.

Keywords: dyslexia, machine learning, early screening, children